

Applying deep learning for adverse pregnancy outcome detection with pre-pregnancy health data

Yu Mu^{1,*}, Kai Feng², Ying Yang³, and Jingyuan Wang²

¹Beihang University, Sino-French Engineer School, 100191 Beijing, China

²Beihang University, Department of Computer Science, 100191 Beijing, China

³National Research Institute for Family Planning, 100081 Beijing, China

Abstract. Adverse pregnancy outcomes can bring enormous losses to both families and the society. Thus, pregnancy outcome prediction stays a crucial research topic as it may help reducing birth defect and improving the quality of population. However, recent advances in adverse pregnancy outcome detection are driven by data collected after mothers having been pregnant. In this situation, if a bad pregnancy outcome is diagnosed, the parents will suffer both physically and emotionally. In this paper, we develop a deep learning algorithm which is able to detect and classify adverse pregnancy outcomes before parents getting pregnant. We train a multi-layer neural network by using a dataset of 75542 couples' multi-dimension pre-pregnancy health data. Our model outperforms some of algorithms in accuracy, recall and F1 score.

1 Introduction

In recent years, adverse pregnancy outcomes have drawn more and more attention across the world. More than 8 million new-born babies have birth defects worldwide every year. In different countries, new-borns with birth defects account for about 3-6% of all new-borns. China is one of the countries with a high rate of neonatal birth defects. According to reports, the birth defect rate is as high as 4-6%, and about 0.8-1.2 million children with defects are born in China every year. On average, a child with a defect is born every 30 seconds. What is even more worrying is that in recent years, the number of birth defects every year has been increasing. Thus, prediction of pregnancy outcome becomes a crucial research topic as it may help reducing birth defect and improving the quality of population. Nowadays machine learning methods are widely used in adverse pregnancy outcome detection [1-2]. However, the process is far from satisfactory. Most recent advances in pregnancy outcome prediction use data collected after mothers get pregnant [3-5]. In this situation, if an adverse pregnancy outcome is diagnosed, the parents will suffer physically, financially and emotionally. Besides, people only focus on a few or a dozen important features for prediction in most studies [6-7]. In this paper, we propose a deep learning model which is able to detect and classify adverse pregnancy outcomes before parents

* Corresponding author: vannessamuyu@buaa.edu.cn

getting pregnant. We train a multi-layer neural network by using a dataset of 75542 couples' multi-dimension pre-pregnancy health data. The dataset includes 308 features. They include much of information such as medical examination, clinical test, history of diseases and drug use, family history of disease, history of pregnancy, demographic characteristics, lifestyle and environment information for both wife and husband, separately. The dataset also includes the true pregnancy outcome of each couple, which is the label for the following prediction. The labels include six types of pregnancy outcomes: normal, premature birth, low birth weight, birth defect, spontaneous abortion and stillbirth. After several observations, our model precisely classifies six classes of pregnancy outcomes and achieves better performance in accuracy, recall and F1 score than some other algorithms do.

The rest of the paper is organized as follows: Sect. 2 presents briefly about the related work. Proposed adverse pregnancy outcome detection and classification model is presented in Sect. 3. Experimental details and results are described in Sect. 4. Finally, we present conclusion and future work in Sect. 5.

2 Related work

Traditional detection of diseases is mainly based on the Cox proportional hazards regression model (Cox model) and logistic regression model. For example, Wang et al. [8] published in 2003 an article using the Cox model to establish a risk prediction model for stroke and death in patients with atrial fibrillation based on the Framingham heart study. The Cox model has been widely used in medical research as a multi-factor regression analysis method in traditional survival analysis and risk prediction.

However, although traditional regression methods have a wide range of applications in disease prediction, these methods still have room for improvement in terms of accuracy and interpretability. In recent years, feature selection and supervised learning modelling methods are increasingly used for disease detection. On one hand, some machine learning methods can improve the interpretability of predictive models, such as decision tree methods [9]. On the other hand, we can use machine learning methods to achieve better predictive performance. Khosla et al. published an article published in SIGKDD in 2010 [10] used feature selection and machine learning methods to predict the incidence of stroke within 5 years. The study used three feature selection methods, including forward feature selection, L1 regularization and conservative mean feature selection. At the time of modelling, support vector machine (SVM) and edge-based censored regression methods were attempted.

For the analysis of electronic medical record data, some studies have also used deep learning methods such as CNN or RNN to establish disease risk prediction models. Cheng et al. [11] developed a CNN network to predict future events based on 4-year EHR data from more than 300,000 patients. Chio et.al [12] was the first to use the RNN-based approach to the prediction of heart failure (HF) to analyse temporal relation before clinical events in electronic medical records.

3 Model

3.1 Problem formulation

The adverse pregnancy outcome detection is a task which takes as input a feature vector $\mathbf{X} = [x_1, \dots, x_m]$, and outputs a predicted label vector $\hat{\mathbf{Y}} = [y_1, \dots, y_n]$, such that each y_i represents one of the six pregnancy outcome classes. For a single sample in the training set, we optimize the cross-entropy objective function

$$\mathcal{L}(\mathbf{X}, \hat{\mathbf{Y}}) = -\frac{1}{n} \sum_{i=1}^n t_i \log p(Y = y_i | \mathbf{X}) \tag{1}$$

where t_i is the true label of the sample, and $p(\cdot)$ is the probability the network assigns to the i -th output taking on the value y_i , i.e.

$$p(Y = y_i | \mathbf{X}) = \frac{\exp(y_i)}{\sum_i \exp(y_i)}, i = 1, \dots, n \tag{2}$$

3.2 Proposed network architecture

The proposed model is shown in fig.1. After the pre-processing is done, the input is divided into several subgroup according to the information catalogues they represented, for wife and husband separately. Then we have 7+6 sub-input blocks (Male have no pregnancy history information). The number of neurons in each sub-input block differs and depends on the number of features included in the catalogue. Then each sub-input block is connected with its own 2 hidden layers and reached pre-output layer as two neurons. Finally, we obtain an output by putting the 13*2 neutrons in pre-output layer into a softmax. Based on this network, we classify the samples to any of the six pregnancy outcome classes.

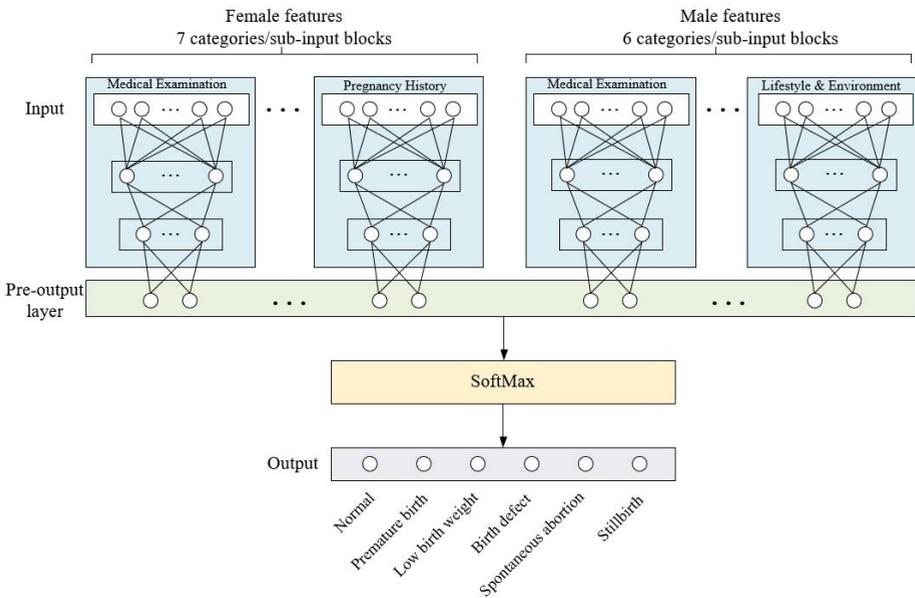


Fig. 1. Network architecture.

4 Experiments

4.1 Dataset

This study is based on the health data derived from the National Free Pre-Pregnancy Check-ups (NFPC), a population-based health survey of reproductive-aged couples who wish to conceive. It was conducted across 31 provinces in China from January 1, 2014 to December 31, 2015.

All the features are collected before the couples getting pregnant and the labels are collected after the pregnancy is finished. We excluded samples with missing information on pregnancy outcome. Finally, 75542 couples/samples were included in the current analysis.

In order to accurate out-of-sample prediction, we divide the data into a training dataset, a validation dataset and a test dataset. We realized 5-fold cross validation on the dataset and the ratio of training-validation-test datasets is 3:1:1. Of these samples, 16243 samples have a normal pregnancy outcome. 15382 samples are labelled premature birth, 12537 samples are labelled low birth weight, 11063 samples are labelled birth defect, 10239 samples are labelled spontaneous abortion and 10078 samples are labelled stillbirth.

We have implemented the proposed deep learning model for adverse pregnancy outcome detection and classification using TensorFlow 0.12.1 and Python 2.7.6 on a Linux x86_64 machine. The base learning rate is set to 0.025 and batch size is 10.

4.2 Results

To the best of our knowledge, our approach is the first one for adverse pregnancy outcome detection and classification using NFPC pre-pregnancy health dataset. Table 1. summarizes the results of accuracy, recall and F1 score for all the six pregnancy outcome classes. The average accuracy is 0.892. The average recall and F1 score achieve to 0.668 and 0.670 respectively.

Table 1. Accuracy, recall and F1 results for all pregnancy outcome classes.

	Accuracy	Recall	F1 Score
Normal	0.860	0.764	0.699
Premature birth	0.889	0.667	0.710
Low birth weight	0.905	0.706	0.713
Birth defect	0.901	0.647	0.658
Spontaneous abortion	0.902	0.615	0.628
Stillbirth	0.897	0.606	0.612
Average	0.892	0.668	0.670

We have also implemented and compared the performance of another two machine learning models, a 5-layer fully-connected neural network and a decision tree [13] with that of our model. The performance comparison is presented in table 2. Our model outperforms these 2 algorithms in average accuracy, recall and F1 score.

Table 2. Comparison of the proposed model, a 5-layer NN and a decision tree.

	our model	5-layer NN	Decision tree
Accuracy	0.892	0.859	0.795
Recall	0.668	0.631	0.618
F1 Score	0.670	0.653	0.624

5 Conclusion

Our adverse pregnancy outcome detection and classification model is practical for the early detection and prevention of birth defect. We have proposed a deep learning algorithm to predict pregnancy outcome using pre-pregnancy health data. We demonstrated the performance of the model on NFPC dataset. Currently, we are still working with more optimizations and interpretability to obtain a better and interpretable result. In the future, we hope that this machine learning method enables more widespread use of adverse pregnancy outcome detection in order to minimize the incidence of adverse pregnancy outcome.

This study was approved by the National Health and Family Planning Commission. Informed consents were obtained from all the NFPC participants. We acknowledge funding support from the State's Key Project of Research and Development Plan (2016YFC1000307).

References

1. A. Koivu, T. Korpimäki, P. Kivelä, T. Pahikkala, M. Sairanen. Evaluation of machine learning algorithms for improved risk assessment for Down's syndrome. *Computers in Biology and Medicine* (2018)
2. Y. Luo, Z. Li, H. Guo, et al. Predicting congenital heart defects: A comparison of three data mining methods. *Plos One*, **12(5)**, e0177811 (2017)
3. M.J. Rooijackers, C.Rabbotti, H.de Lau, S.G.Oei, J.W.M Bergmann's, M. Mischi. Feasibility study of a new Method for low complexity fetal movement detection from abdominal ECG recordings. *Journal of Biomedical and Health Informatics, IEEE*, pp.2168-2194 (2015)
4. B.N. Lakshmi, T.S. Indumathi, R. Nandini. A hybrid approach for prediction based health monitor in pregnant women. *ICETEST, Elsevier* (2016)
5. H. Sahin, A. Subasi. Classification of the cardiotocogram data for anticipation of fetal risks using machine learning techniques. *Applied Soft Computing*, **33(C)**, 231-238(2015)
6. H. Ocak. A medical decision support system based on support vector machines and the genetic algorithm for the evaluation of fetal well-being. *Journal of Medical Systems*, **37(2)**, 9913 (2013)
7. M.L. Huang, Y.Y. Hsu. Fetal distress prediction using discriminant analysis, decision tree, and artificial neural network. *Journal of Biomedical Science & Engineering*, 05(9) (2012)
8. T.J. Wang, J.M. Massaro, D. Levy, et al. A risk score for predicting stroke or death in individuals with new-onset atrial fibrillation in the community: the Framingham Heart Study. *Jama*, **12(6)**,1049-1056 (2003)
9. G.C. Fonarow, K.F. Adams, W.T. Abraham, C.W. Yancy, W.J. Boscardin. Risk stratification for in-hospital mortality in acutely decompensated heart failure: classification and regression tree analysis. *Jama*, **293(20)**, 2467 (2005)
10. A. Khosla, Y. Cao, C.Y. Lin, et al. An integrated machine learning approach to stroke prediction. *ACM SIGKDD. ACM*, 183-192 (2010)
11. Y. Cheng, F. Wang, P. Zhang, et al. Risk Prediction with Electronic Health Records: A Deep Learning Approach. *Siam International Conference on Data Mining*, 432-440 (2016)

12. E. Choi, A. Schuetz, W.F. Stewart, et al. Using recurrent neural network models for early detection of heart failure onset. *Jamia*, **24(2)**, 361 (2016)
13. J.R. Quinlan. Induction on decision tree. *Machine Learning*, **1(1)**, 81-106 (1986)